Problem 1 Part 2:

model: $g_{w,b}(x) = wx + b$

$$\mathcal{E}(w,b) = \frac{1}{2N}\sum_{i=1}^{N}\left(g_{w,b}(x_i) - t_i\right)^2$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\left(w x_i + b - t_i\right)^2$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\left(x_i^2 w^2 + b^2 + 2x_i b w - 2t_i x_i w - 2t_i b + t_i^2\right)$$

$A_i(x_i, t_i) = x_i^2$ $\quad B_i(x_i, t_i) = 1$ $\quad C_i(x_i, t_i) = 2x_i$

$D_i(x_i, t_i) = -2x_i \cdot t_i$ $\quad E_i(x_i, t_i) = -2t_i$ $\quad F_i(x_i, t_i) = t_i^2$

Problem 7 Part 3:

$$\frac{1}{2N}\sum_{i=1}^{N}\left(x_i^2 w^2 + b^2 + 2x_i bw - 2t_i x_i w - 2t_i b + t_i^2\right)$$

$A_i(x_i, t_i) = x_i^2$ $\quad$ $B_i(x_i, t_i) = 1$ $\quad$ $C_i(x_i, t_i) = 2x_i$

$D_i(x_i, t_i) = -2x_i \cdot t_i$ $\quad$ $E_i(x_i, t_i) = -2t_i$ $\quad$ $F_i(x_i, t_i) = t_i^2$

$$\mathcal{E} = \frac{1}{2N}\left(A w^2 + B b^2 + C wb + D w + E b + F\right)$$

Find $\frac{\partial \mathcal{E}}{\partial w}, \frac{\partial \mathcal{E}}{\partial b}$

$$\frac{\partial \mathcal{E}}{\partial w} = \frac{2Aw + Cb + D}{2N}, \quad \frac{\partial \mathcal{E}}{\partial w} = \frac{2Bb + Cw + E}{2N}$$

Set $= 0$

$$\frac{\partial \mathcal{E}}{\partial w} = 0 = \frac{2Aw + Cb + D}{2N} \Rightarrow 2Aw + Cb + D = 0 \Rightarrow w = \frac{-Cb - D}{2A}$$

$$\frac{\partial \mathcal{E}}{\partial b} = 0 = \frac{2Bb + Cw + E}{2N} \Rightarrow 2Bb + Cw + E = 0 \Rightarrow b = \frac{-Cw - E}{2B}$$

$$w = \frac{EC - 2DB}{4AB - C^2}$$

$$b = \frac{CD - 2EA}{4AB - C^2}$$

Problem 2 Part 1:

$$\vec{x}_i = \begin{bmatrix} x_i & 1 \end{bmatrix}$$

$$\vec{w} \Rightarrow \begin{bmatrix} w \\ b \end{bmatrix}$$

$$\therefore g_w(\vec{x}_i) = \vec{x}_i \vec{w} = \begin{bmatrix} x_i & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = x_i \cdot w + 1 \cdot b = w x_i + b = g_{w,b}(x_i)$$

Therefore the Vectorized Version is the same as the Scaler Version.

Problem 2 Part 2:

$$\nabla_{\vec{w}} \| X\vec{w} - \vec{t} \|^2 \qquad X = \begin{bmatrix} x^{(1)} & 1 \\ x^{(2)} & 1 \\ \vdots & \vdots \\ x^{(N)} & 1 \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} w \\ b \end{bmatrix} \quad \vec{t} = [t^{(1)} \; t^{(2)} \cdots t^{(N)}]$$

$$\nabla_{\vec{w}} = \begin{bmatrix} \frac{\partial}{\partial w} \\ \frac{\partial}{\partial b} \end{bmatrix}$$

Solve:

$$X\vec{w} = \begin{bmatrix} w \cdot x^{(1)} + 1 \cdot b \\ w \cdot x^{(2)} + 1 \cdot b \\ \vdots \quad\quad \vdots \\ w \cdot x^{(N)} + 1 \cdot b \end{bmatrix} \Rightarrow X\vec{w} - \vec{t} = \begin{bmatrix} w \cdot x^{(1)} + b - t^{(1)} \\ w \cdot x^{(2)} + b - t^{(2)} \\ \vdots \quad\quad \vdots \\ w \cdot x^{(N)} + b - t^{(N)} \end{bmatrix}$$

$$\Rightarrow \| X\vec{w} - \vec{t} \|^2 = \left[\left(\sum_{i=1}^{N} (w x^{(i)} + b - t^{(i)})^2 \right)^{\frac{1}{2}} \right]^2 = \sum_{i=1}^{N} (w x^{(i)} + b - t^{(i)})^2$$

$$\Rightarrow \nabla_{\vec{w}} \| X\vec{w} - \vec{t} \|^2 = \begin{bmatrix} \frac{\partial}{\partial w} \\ \frac{\partial}{\partial b} \end{bmatrix} \sum_{i=1}^{N} \underbrace{(w x^{(i)} + b - t^{(i)})^2}$$

From Problem 1 $\Rightarrow Aw^2 + Bb^2 + Cwb + Dw + Eb + F$

$$= \begin{bmatrix} \frac{\partial \sum (w x^{(i)} + b - t^{(i)})^2}{\partial w} \\ \frac{\partial \sum (w x^{(i)} + b - t^{(i)})^2}{\partial b} \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{\partial\, Aw^2 + Bb^2 + Cwb + Dw + Eb + F}{\partial w} \\ \frac{\partial\, Aw^2 + Bb^2 + Cwb + Dw + Eb + F}{\partial b} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 2Aw + Cb + D \\ 2Bb + Cw + E \end{bmatrix} = \nabla_{\vec{w}} \| X\vec{w} - \vec{t} \|^2$$

Problem 2 Part 2:

$\nabla_{\vec{w}} \| X\vec{w} - \vec{t} \|^2$

Solve: $\nabla_{\vec{w}} \| X\vec{w} - \vec{t} \|^2 \Rightarrow \nabla_{\vec{w}} \left[(X\vec{w} - \vec{t})^T (X\vec{w} - \vec{t}) \right]$

$\Rightarrow \nabla_{\vec{w}} \left[(X^T \vec{w}^T - \vec{t}^T)(X^T \vec{w} - \vec{t}) \right] \Rightarrow \nabla_{\vec{w}} \left[\vec{w}^T X^T X \vec{w} - \vec{w}^T X^T \vec{t} - \vec{t}^T X \vec{w} + \vec{t}^T \vec{t} \right]$

$\Rightarrow \nabla_{\vec{w}} (\vec{w}^T X^T X \vec{w}) - \nabla_{\vec{w}} (\vec{w}^T X^T \vec{t}) - \nabla_{\vec{w}} (\vec{t}^T X \vec{w}) + \nabla_{\vec{w}} (\vec{t}^T \vec{t})$

$\Rightarrow (X^T X + (X^T X)^T)\vec{w} - X^T \vec{t} - X^T \vec{t} + 0$

$\Rightarrow 2X^T X \vec{w} - 2X^T \vec{t}$

Problem 2 Part 3:

$$2X^TX\vec{w}^* - 2X^T\vec{t} = 0$$

To minimize least Squares $\Rightarrow$ $\underbrace{\nabla_{\vec{w}} \| X\vec{w} - \vec{t} \|^2}_{\text{From Problem 2 Part 2} = 2X^TX\vec{w} - 2X^T\vec{t}} = 0$

Therefore any $\vec{w}$ that minimizes the least Squares must Satisfy

$$2X^TX\vec{w}^* - 2X^T\vec{t} = 0$$

Problem 2 Part 4:

assuming $(X^T X)(X^T X)^{-1} = I$ (identity):

$$\Rightarrow 2X^T X \vec{w}^* - 2X^T \vec{t} = 0 \Rightarrow X^T X \vec{w}^* = X^T \vec{t}$$

$$\Rightarrow (X^T X)^{-1} X^T X \vec{w}^* = (X^T X)^{-1} X^T \vec{t}$$

$$\Rightarrow \vec{w}^* = (X^T X)^{-1} X^T \vec{t}$$

Problem 3 Part 1:

$$A = \sum_{i=1}^{N} \vec{x}^{(i)} \vec{x}^{(i)T}$$

$$= \sum_{i=1}^{N} \underbrace{\sum_{j=1}^{d} x_j^{(i)} \cdot x^{(i)T}}_{\Downarrow}$$

$$\begin{bmatrix} x_1^{(i)2} & x_1^{(i)} x_2^{(i)} & \cdots & x_1^{(i)} x_d^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ x_d^{(i)} x_1^{(i)} & x_d^{(i)} x_2^{(i)} & \cdots & x_d^{(i)2} \end{bmatrix}$$

$d \times d$

$$\therefore A = \left[a_{ij} = \sum_{k=1}^{N} x_i^{(k)} x_j^{(k)} \,\middle|\, i,j \in 1, \dots, d \right]$$

i.e

$$A = \begin{bmatrix} x_1^{(1)2} & x_1^{(1)} x_2^{(1)} & \cdots & x_1^{(1)} x_d^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_d^{(1)} x_1^{(1)} & x_d^{(1)} x_2^{(1)} & \cdots & x_d^{(1)2} \end{bmatrix} + \cdots + \begin{bmatrix} x_1^{(N)2} & x_1^{(N)} x_2^{(N)} & \cdots & x_1^{(N)} x_d^{(N)} \\ \vdots & \vdots & \ddots & \vdots \\ x_d^{(N)} x_1^{(N)} & x_d^{(N)} x_2^{(N)} & \cdots & x_d^{(N)2} \end{bmatrix}$$